Science Data Processing Workshop 2000

# ICESat Science Data Systems

Nov 7-8, 2000

David Hancock, NASA/GSFC Code 972

Anita Brenner, Raytheon/ITSS

Mark Sherman, Raytheon/ITSS

hancock@osb.wff.nasa.gov

757 824 1238

Science Data Processing Workshop 2000

# ICESat- **I**ce, **C**loud, and land **E**levation **Sat**ellite

- Single Instrument – Geoscience Laser Altimeter System (GLAS)
- Mission designed to measure
  - ice-sheet topography and associated temporal changes
  - cloud and atmospheric properties
  - land and water along-track topography.
- Launch December, 2001

Science Data Processing Workshop 2000

# ICESat Science Data Software Components Presented

- ICESat Science Investigator-led Processing System (I-SIPS) Software
  - Schedule and Data Management Subsystem (SDMS)- Processing environment to control job flow, data distribution, and archiving
  - GLAS Science Algorithm Software (GSAS) – Creates GLAS standard products (controls which products are created and implements ATBD )
- GLAS Science Computing Facility (SCF) Software
  - Visualization
  - Geographic and Temporal Subsetting (GATS)

Science Data Processing Workshop 2000

# ICESat SOFTWARE DEVELOPMENT DESCRIPTION

- Scheduling and Data Management subsystem (SDMS) being coded at GSFC under I-SIPS proposal
- GLAS science algorithm software being coded at GSFC under Science Team direction
  - Based on ATBDs
  - Being implemented by one set of developers
  - Under configuration management (using ClearCase)
  - Software designed to handle processing, partial processing, and reprocessing
- GLAS SCF Software being coded under code 971 science team member

Science Data Processing Workshop 2000

# I-SIPS PROCESSING BASIC REQUIREMENTS

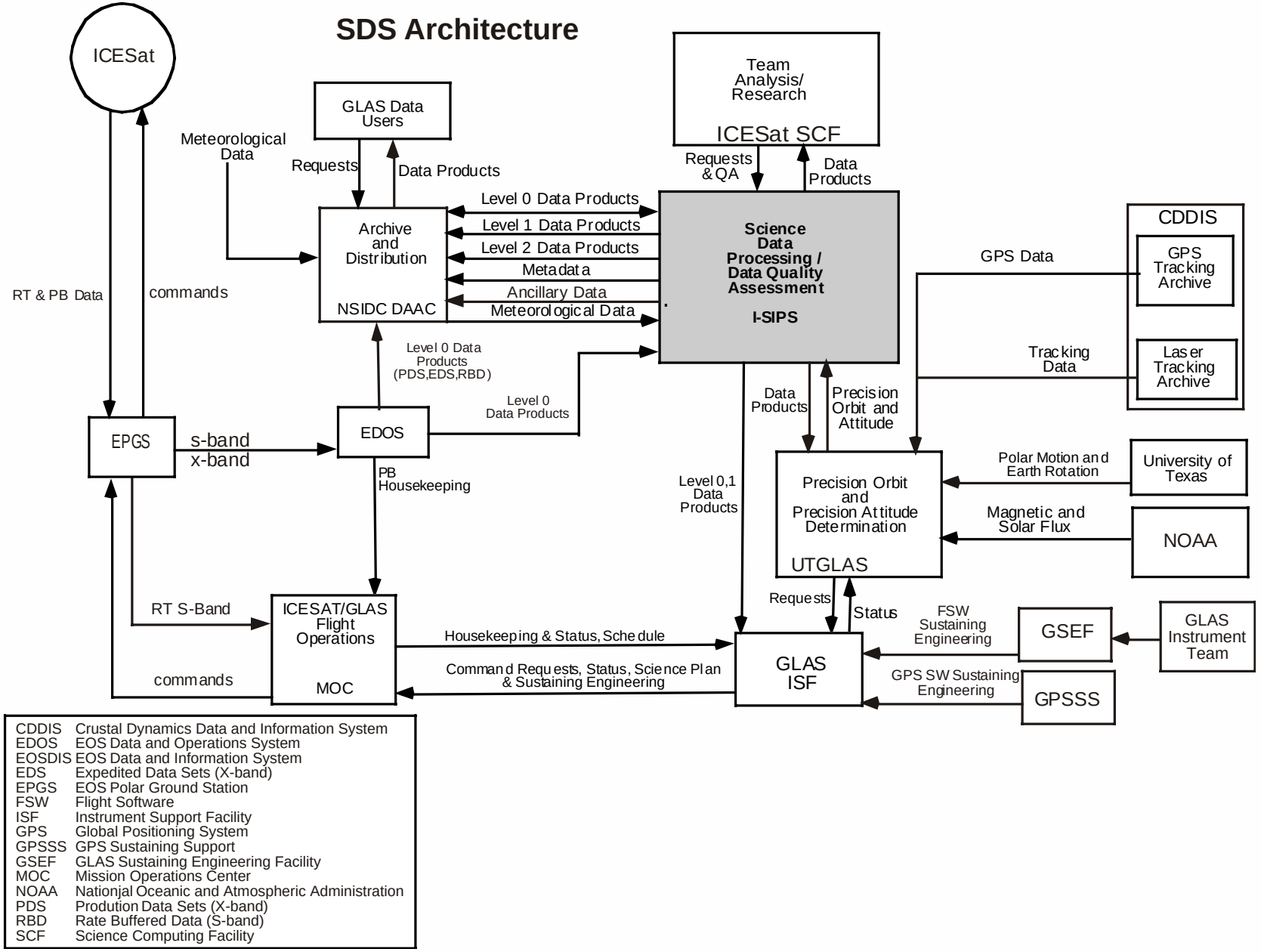
- Process 24 hours of GLAS instrument data into standard data products within 4 hours of receipt of all required inputs
- Ability to distribute to the Science Team Level 1 and Level 2 data products within 24 hours of receipt of Level 0 data (uses predict ancillary data)
- Distribute fully processed Level 1 and Level 2 data products to NSIDC within 14 days of receipt of Level 0 data (after becoming operational and assuming proper funding)
- Support reprocessing requirements without delaying regular processing assuming proper funding

Science Data Processing Workshop 2000

# I-SIPS DATA ARCHIVING REQUIREMENTS and OPERATIONS

- Data Archiving
  - Archives internal data products for internal I-SIPS use, to End-of-Mission
  - I-SIPS archives a log of products delivered
  - I-SIPS does not perform permanent archive
- Operations
  - Autonomous Operation 7 days/week, 24 hours
  - Normal Manned Operation is 5 days/week, 12 hours
  - Available on-call
  - Initial calibration period TBD (as many as required)

# I-SIPS ARCHITECTURAL DIAGRAM




| | |
|---|---|
| CDDIS | Crustal Dynamics Data and Information System |
| EDOS | EOS Data and Operations System |
| EOSDIS | EOS Data and Information System |
| EDS | Expedited Data Sets (X-band) |
| EPGS | EOS Polar Ground Station |
| FSW | Flight Software |
| ISF | Instrument Support Facility |
| GPS | Global Positioning System |
| GPSSS | GPS Sustaining Support |
| GSEF | GLAS Sustaining Engineering Facility |
| MOC | Mission Operations Center |
| NOAA | Nationjal Oceanic and Atmospheric Administration |
| PDS | Prodution Data Sets (X-band) |
| RBD | Rate Buffered Data (S-band) |
| SCF | Science Computing Facility |

Science Data Processing Workshop 2000

# Scheduling and Data Management System (SDMS)

- The SDMS is software developed for ISIPS that:
  - Ingests data from external systems
  - Automatically applies science processing to ingested data to produce new products
  - Archives products into robotic digital libraries
  - Tracks all products ever created
  - Reprocesses data when new versions of science algorithms or new supporting data sets are released
  - Automatically distributes data to the NSIDC DAAC for further distribution to end users
  - Has high potential reuse in other science missions

# High Level SDMS Requirements for ISIPS

- Scheduling Requirements
  - 24x7 operation with minimal operator intervention
  - Automatically start processing when needed inputs available
  - Provide effective error recovery procedures
  - Allow ISIPS staff to monitor, control, prioritize workload
  - Manage system resources (Disk, CPU) for best throughput
- Data Management Requirements
  - Keep copy of all data products produced
  - Maintain product metadata in searchable database
  - Retain processing history for long-term problem analyses

Science Data Processing Workshop 2000

# Additional SDMS Goals

- Scheduling
  - Plug-in of GLAS science processing and supporting utilities with minimal modifications to the existing code base
  - No forced use of a specific toolkit *(freedom to choose what makes sense)*
  - Integrated scheduling across multiple computer platforms (primary and backup)
- Data Management
  - Keep two copies of all data products for additional protection against data loss

Science Data Processing Workshop 2000

# We Started with V0 DAAC

- The V0 DAAC project had Scheduling and Data Management tools similar to what was needed
- An agreement was formed between the two projects where:
  - the ISIPS project would use selected portions of the V0 S/W
  - the ISIPS project would return improvements to the DAAC

# Key Changes Made to V0 DAAC Scheduling

- The V0 DAAC provided a scheduling system that when instructed to run a job would:
  - Execute the processes of jobs, watching for errors
  - Manage system resources such as disk and CPU
  - Provide a good set of displays for visualizing system status
- We also needed to develop a planning capability that:
  - Decides when a job has everything it needs to run
  - Builds job control files that dictate what a job does
  - Enters job onto a prioritized queue of work to be done
  - Dispatches jobs to the scheduling system as needed
  - Spreads jobs across multiple computer systems

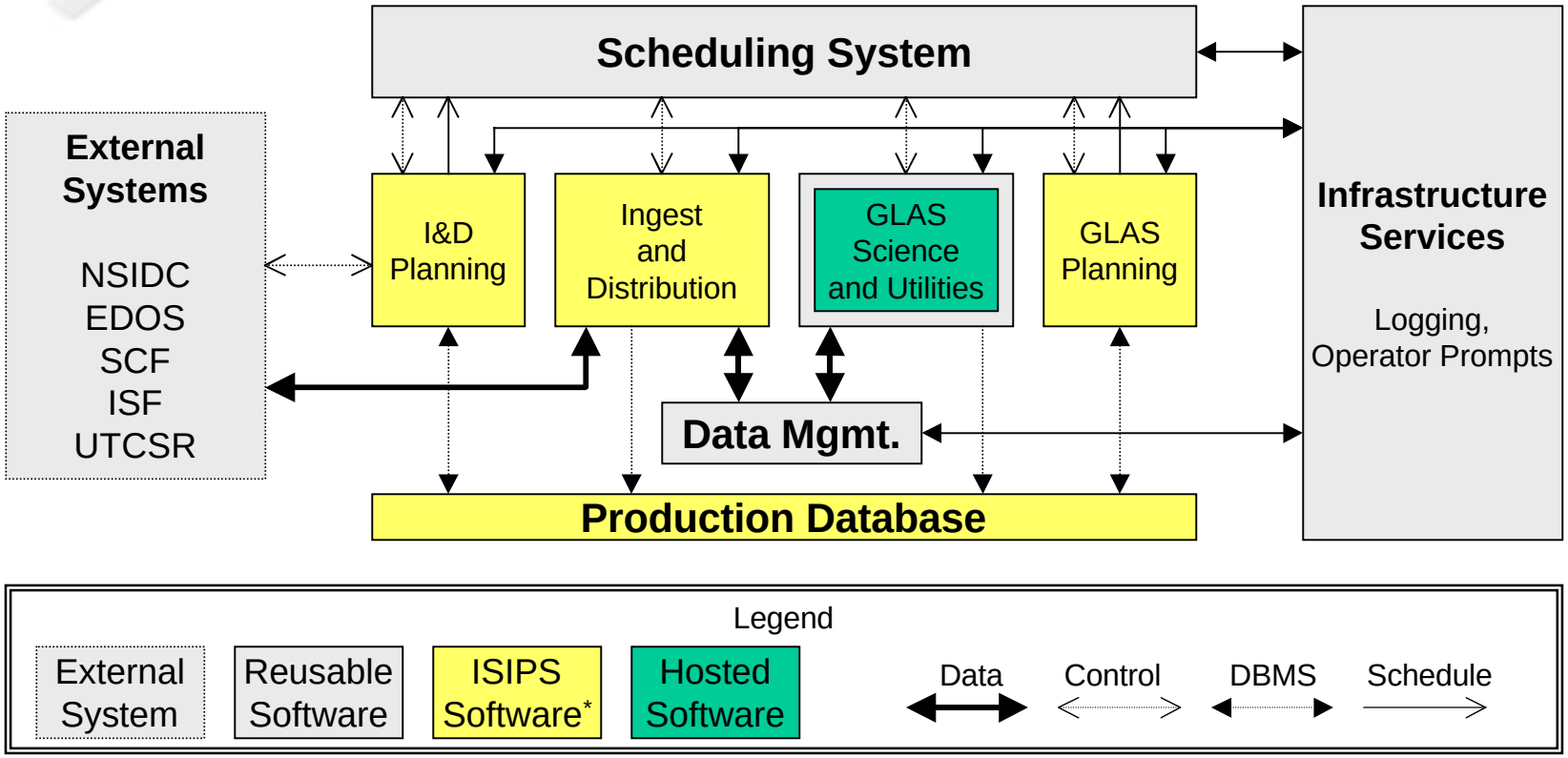
Science Data Processing Workshop 2000

# Key Changes Made to V0 DAAC Data Management

- The ARCHER robotic file management software was used rehosted from the DAAC SGI to the ISIPS HP
- We wrote a new Data Server (cache management) system using the V0 DAAC Distribution Cache Manager (DCM) as a model:
  - **The V0 DAAC DCM only supported distribution of files from cache**
  - **V0 DAAC did ingest through a sophisticated process that was deemed unnecessarily complex for ISIPS**
- The DAAC provided a good basis for the Production Database:
  - **We removed a number of tables that were used for DAAC-specific functions such as distribution of products to worldwide users**
  - **We (partly) reused many of the tables for tracking products and metadata**
  - **We added support for production planning and status to implement processing rules in support of automated science processing**

Science Data Processing Workshop 2000

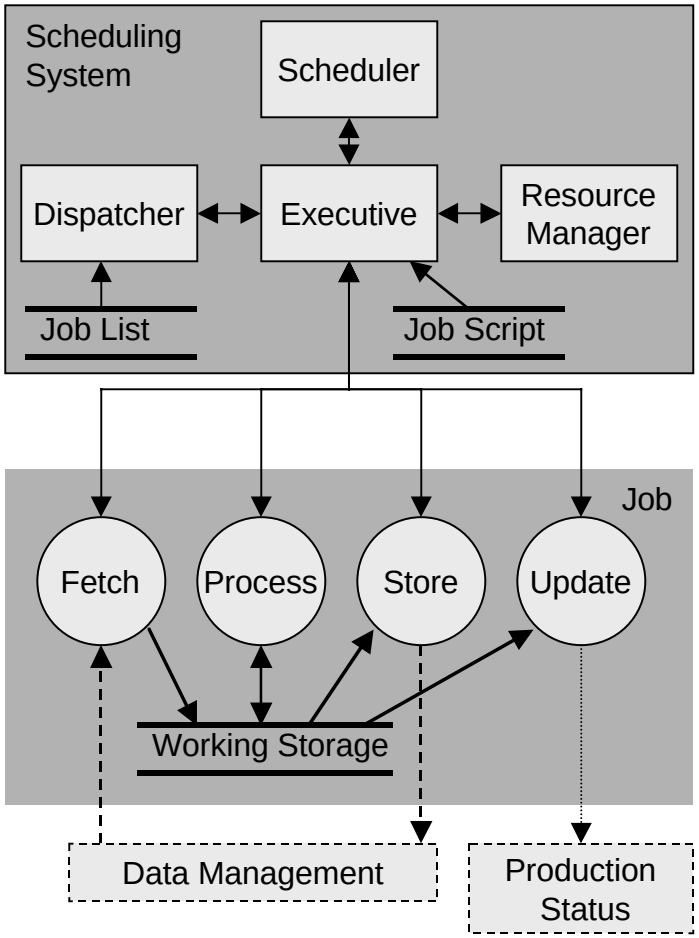
# SDMS Functional Architecture




* Even ISIPS software will have some reuse in other systems

# Scheduling System Details

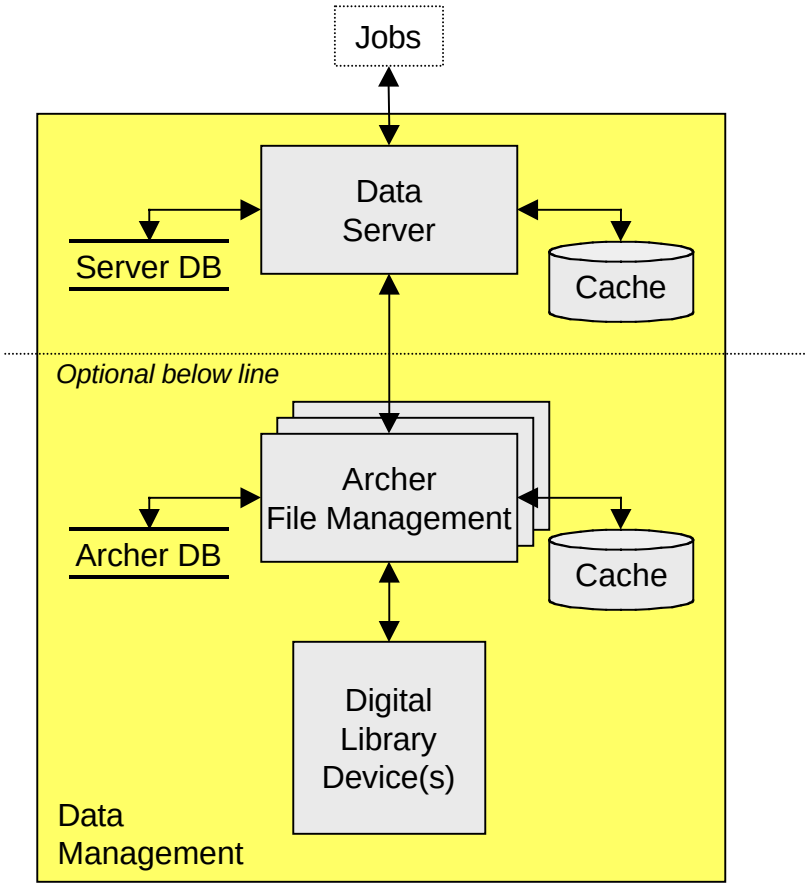
- **Scheduling runs jobs and steps**
  - **Steps are simply Unix processes**
  - **Jobs are steps that run together**
  - **Steps share a working directory**
  - **Scripted steps are in series or parallel**
  - **Multiple jobs can be run at once**
- **Scheduling has four components (most with GUI displays)**
  - **Dispatcher starts and tracks jobs**
  - **Executive controls active steps**
  - **Scheduler controls active jobs**
  - **Resource Manager allocates resources**
- **Typical processing job has four steps**
  - **Fetch to get data from Data Mgmt.**
  - **Process the data (science application)**
  - **Store results in Data Management**
  - **Update production status in DBMS**



Science Data Processing Workshop 2000

# Data Management Details

- Data Management is a large repository for product files and other important data files
- Files can be held on disk, in near-line storage, or off-line
- Files are fetched/stored by name, location is transparent to client job
- Data server is high performance caching system
- ARCHER is optional archive file management system that moves files to/from digital library storage
- Use ARCHER system (or other HSM) only if more data is to be held than disk space is available
- Multiple archive systems like ARCHER may be used if desired



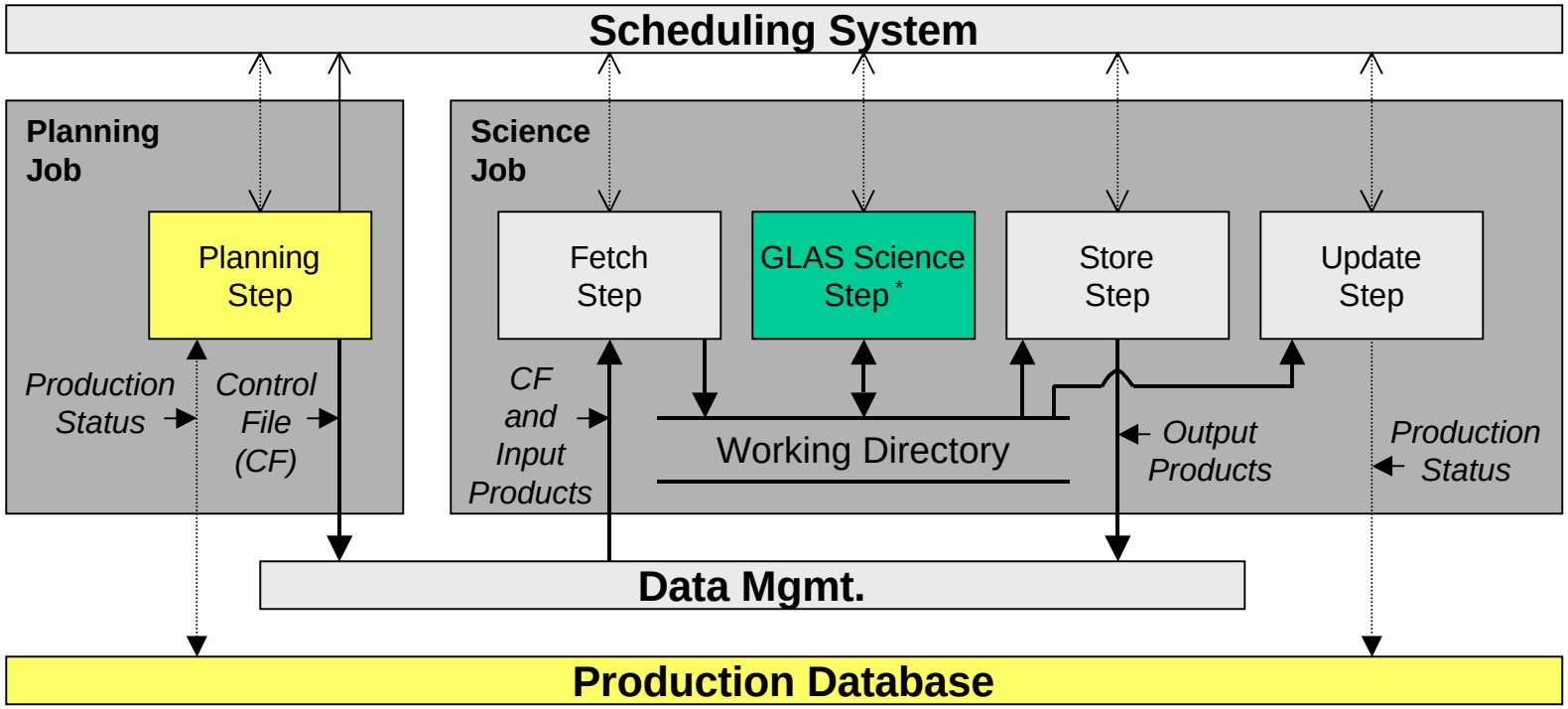
Science Data Processing Workshop 2000

# Production Database

- Foundation for our data-driven design
- Product Tables for Data Management
  - Track product status
  - Retain metadata for products
- Planning Tables for Scheduling:
  - Contains rules specifying what input products are required to run what GLAS utilities to produce what output products
  - Different product and software versions use different sets of production rules
- The status in the Product Tables is processed through the rules in the Planning Tables to determine when a processing job should be run
- From the Product tables, it is possible to work backwards through the Planning tables to determine what specific inputs and software utilities were used to produce any specific product

Science Data Processing Workshop 2000

# Science Processing Planning and Execution Model




* GLAS Science is a Unix process, accepts command line from executive, does all I/O to working directory, returns 0 for success, non-zero for failure. Reads control file (as do the other steps) to determine what to do. These are the *only* interface requirements to run science processing in SDMS (i.e., no toolkit interface required).

Science Data Processing Workshop 2000

# Potential Reuse of SDMS

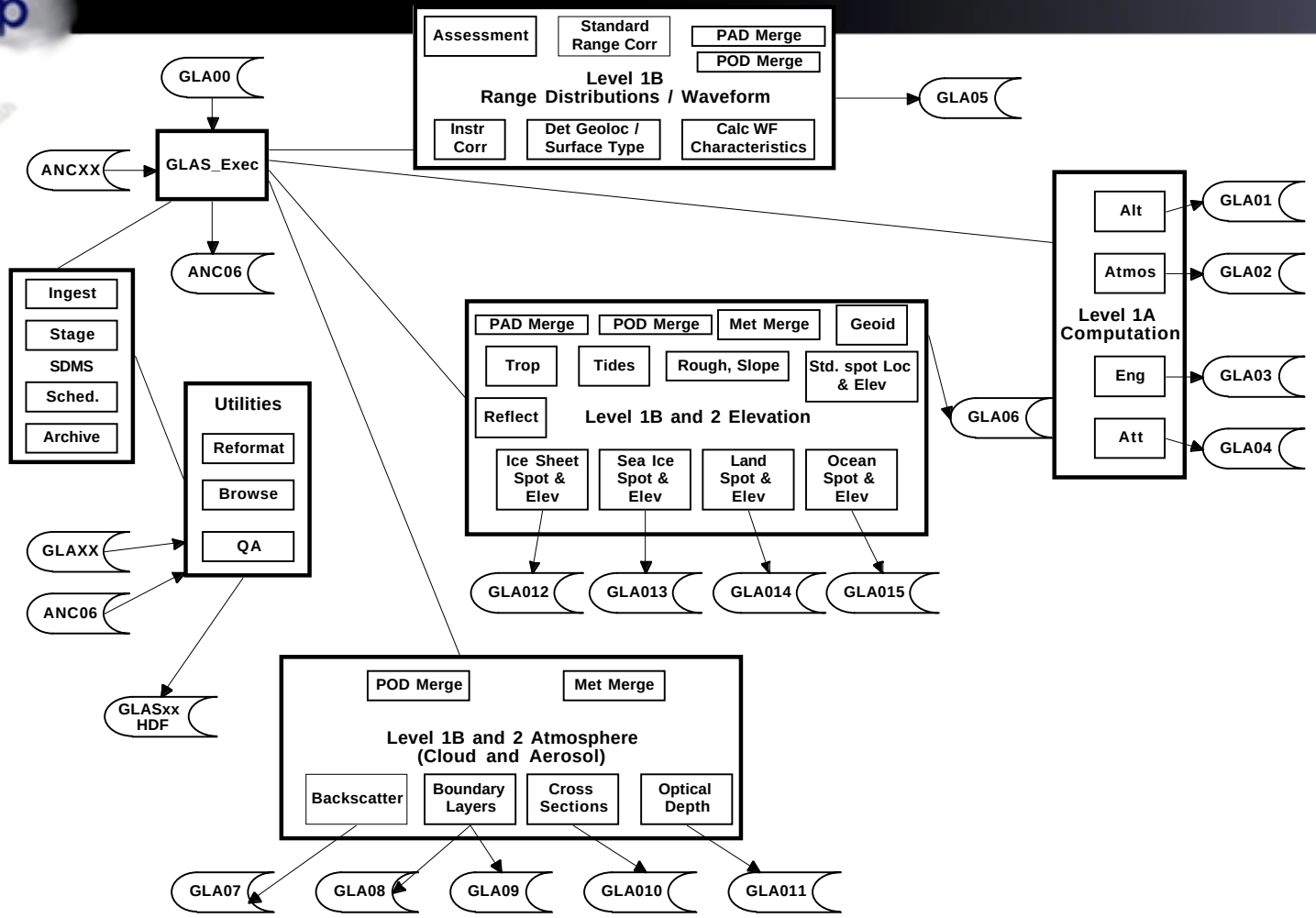
- We started with a reusable suite of software from the GSFC V0 DAAC
- We have produced an even more reusable suite of software for science processing
  - **Scheduling system can be reused**
  - **Data Management system can be reused**
  - **Infrastructure Services can be reused**
  - **Selected interface components can be reused (e.g., DAAC, EDOS)**
  - **Production Planning and Status needs to be developed on a per project basis (but model can be reused)**
  - **Science software plugs in very easily**

# GSAS - GLAS Science Algorithm Software

- Includes all Science Algorithms based on ATBDs as submitted to ESDIS.
- GLAS_Exec can create all products at once, or in steps. For example:
  - Level 1
  - Level 1B and 2 Atmospheric
  - Level 1B and 2 Elevation
- Modularity - orchestrated by GLAS_Exec
  - Selective calls can be made to subsystems or Processes within subsystems.
  - Allows selective processing and reprocessing.
- Each Subsystem is implemented as a Shared Library.
- Each Subsystem computes statistics for Quality Assurance.
- Error Handling is hierarchical. Subsystems will not stop the system directly.

Science Data Processing Workshop 2000

# GSAS SOFWARE TOP LEVEL DECOMPOSITION




I-SIPS SOFTWARE TOP LEVEL DECOMPOSITION


# GSAS SOFTWARE DEVELOPMENT PROCESS LIFE CYCLE

- **Requirements Phase**
  - **Concept and Initiation**
  - **Requirements Development**
- **Design Phase**
  - **Prototyping**
  - **Architectural Design**
- **Implementation and Testing Phase**
  - **Implementation/Coordination**
  - **Integration and Test**
- **Acceptance and Delivery Phase**
  - **Acceptance**
  - **Delivery**
- **Sustaining Engineering and Operations Phase**
  - **Operations**
  - **Maintenance**


# GSAS SOFTWARE DEVELOPMENT PROCESS PROCEDURES

- Design Review
- PDL Review
- Code Review
- Unit Test Reviews
- Testing
  - Unit Testing
  - Integration Testing
  - Acceptance Testing

# GSAS DOCUMENTATION TREE

| GLAS Standard Data Software | |
|---|---|
| Management Plan Volume | GLAS Science Software Management Plan |
| | GLAS Science Data Management Plan |
| Product Specification Volume | GLAS Science Software Requirements |
| | GLAS Level 0 Instrument Data Product Specification |
| | GLAS Standard Data Products Specification – Level 1 |
| | GLAS Standard Data Products Specification – Level 2 |
| | GLAS Science Software Architectural Design |
| | GLAS Science Software Detailed Design |
| | GLAS Science Software User's Guide/Operational Procedures Manual |
| | GLAS Science Software Version Description |
| Assurance and Test Procedures Volume | GLAS Science Software Assurance and Test Procedures |
| Management, Engineering, and Assurance Reports | GLAS Science Software Performance/Status Report |
| | GLAS Science Software Discrepancy Reports |
| | GLAS Science Software Engineering Change Proposal |
| | GLAS Science Software Test Report |

[based on the NASA Software Documentation Standard – Software Engineering Program, NASA-STD-2100-91, July 19, 1991]

Science Data Processing Workshop 2000

# Geographic and Temporal Subsetting (GATS) at GLAS SCF

- Requirements
  - Allow user to create custom data sets in GLAS standard data format for any of the GLAS level 1 and 2 granules, GLA01 – GLA15 for a specific time and/or region (defined by latitude/longitude rectangle)
- Based on dividing the world into geographic bins and setting up a data base management system that defines what geographic bins each granule traverses and specific records within each granule that traverse individual bins

Science Data Processing Workshop 2000

# GLAS requirements that influence subsetting

- Mission
  - Mission designed for a 183-day repeat track orbit
  - GLAS is a nadir-looking instrument that samples the atmosphere and ground 40 times/sec with a 65m diameter footprint every 150 m along the ground track
  - The instrument will regularly be pointed to targets of opportunity ( n times/day) that can be 56km away from nadir for 5 deg off-pointing – therefore must use actual ground location not reference orbit for subsetting.
- Data Set
  - Granules must be record accessible – i.e. do not need to read sequentially through file to access a specific record - direct access files

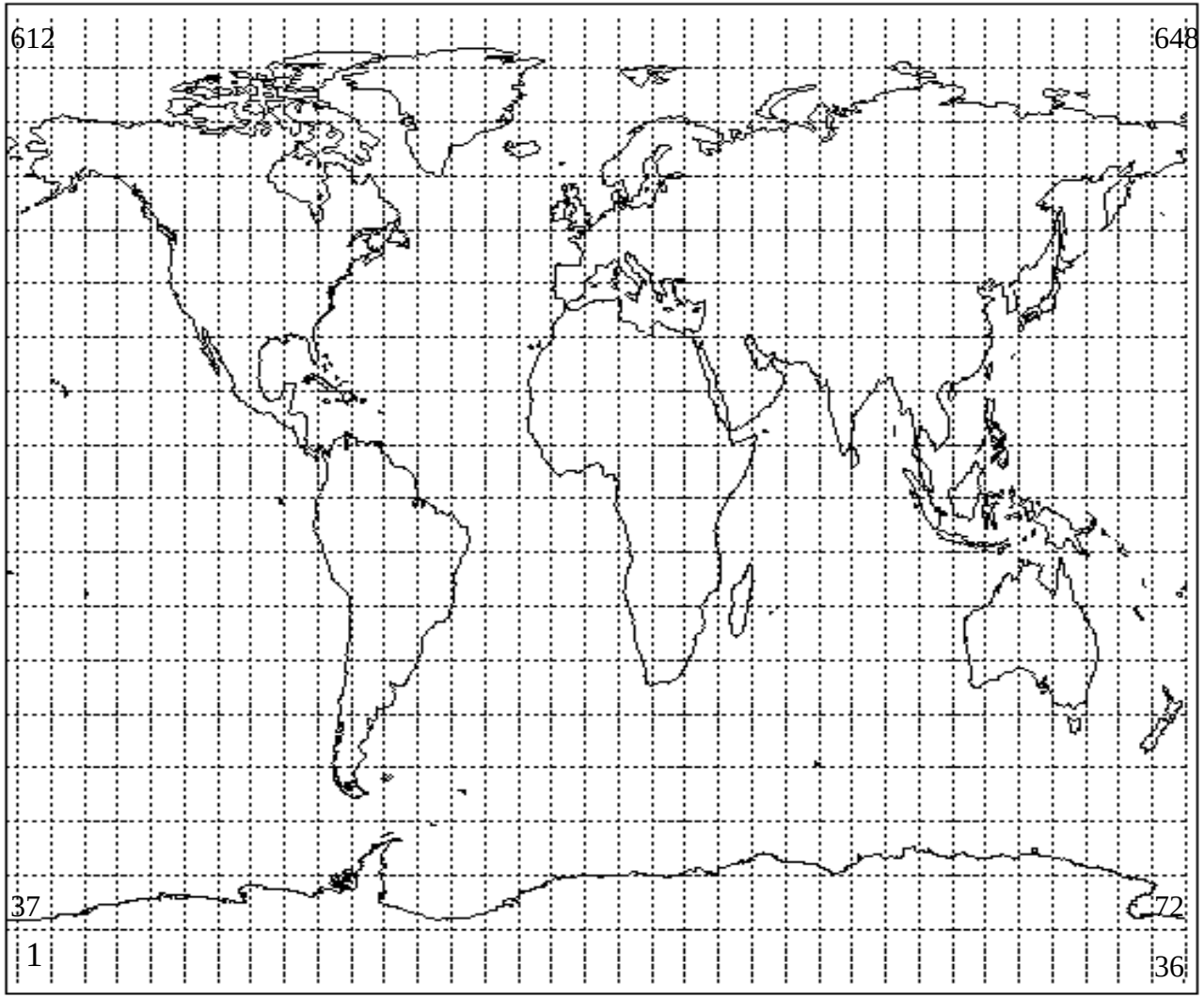
Science Data Processing Workshop 2000

# Geographic and Temporal Subsetting (GATS) at GLAS SCF for level 1 and 2 Granules

- GLAS Products - Fifteen level 1 and 2 products, GLA01-GLA15, that are written in chronological order and vary in granule size from _ revolution to 14 revolutions of data.
- GATS –Provides subset of granules based on requested products, geographic region and time selection.

# Georeference bin Configuration

- Define a georeference bin configuration that divides the world into a set of geographic bins.
  - defining the latitude and longitude of the Southwest corner
  - Define height( in deg latitude)  and width (in degrees longitude) of each row of longitude
  - Given – rectangular region in latitude and longitude
  - Output - Algebraically calculate bin #’s from definition of georeference bin configuration of all bins in the region




Sample georeference bin configuration – southwest latitude/longitude – -90.0, -180.0
18 rows, each with 36 divisions - northeast latitude/longitude - +90.0, 180.0

Science Data Processing Workshop 2000

# GATS-How it works

- Each granule processed to provide start/stop time and start record number and number of consecutive records for each bin that contains data
- This information and granule name is entered in bin database
- Geographical and/or temporal request is made to database to provide direct pointers to all data that match request
- Direct access reads are performed to create the subset data

Science Data Processing Workshop 2000

# Potential Reuse of GATS

- Any project that wants to provide search and retrieval of data by geographical and temporal searches
- Concept is simple, easy to implement, and provides fast subset information
- Basic building blocks can be easily reused
- Details of processing of actual granules are specific to ICESat (granules are binary), but concept is transferable
- Granules stored for direct access take minimum processing to subset